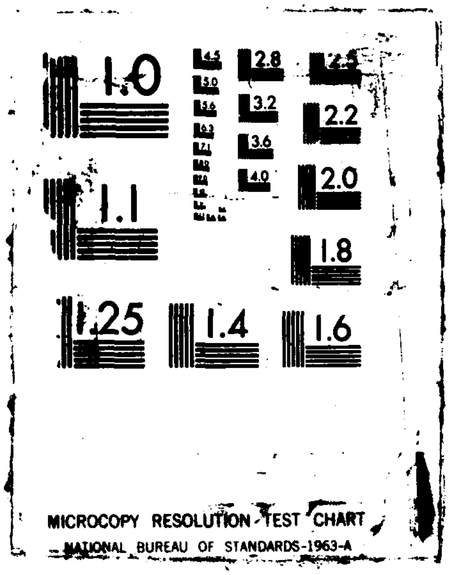
AD-A186 535 AN APL FUNCTION FOR BIVARIATE NORMAL PROBABILITIES(U) 1/1
NAVAL POSTGRADUATE SCHOOL MONTEREY CA T JAYACHANDRAN
AUG 87 NPS53-87-004

UNCLASSIFIED F/G 12/3 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

NPS-53-87-004

NAVAL POSTGRADUATE SCHOOL

Monterey, California

AN APL FUNCTION FOR BIVARIATE NORMAL PROBABILITIES

Toke Jayachandran

Technical Report for Period
October 1986 - July 1987

Approved for public release; distribution unlimited

Prepared for:
Naval Postgraduate School
Monterey, CA 93943-5000

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943

R. C. AUSTIN
Rear Admiral, U.S. Navy
Superintendent

K.T. MARSHALL
Acting Provost

Reproduction of all or part of this report is authorized.

This report was prepared by:

TOKE JAYACHANDRAN
Professor of Mathematics

Reviewed by:

HAROLD M. FREDRICKSEN
Chairman
Department of Mathematics

JAMES M. FREMGEN
Acting Dean of Information and
Policy Science

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A186 535

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NPS-53-87-004	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) An APL Function For Bivariate Normal Probabilities		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Toke Jayachandran		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93943		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93943		12. REPORT DATE August 1987
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

bivariate normal
cumulative distribution function
APL language
microcomputer

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An APL function to compute cumulative probabilities for a standard bivariate normal distribution is presented. The function can be run on an IBM-PC/AT compatible microcomputer such as the Zenith Z-248, as well as on the IBM 3033 mainframe computer.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AN APL FUNCTION FOR BIVARIATE NORMAL PROBABILITIES

AN APL FUNCTION FOR BIVARIATE NORMAL PROBABILITIES

INTRODUCTION: Bivariate normal distributions have many applications such as in combat modelling, weapons systems effectiveness studies and weather prediction problems; several other applications are discussed in [4]. It is well known that probability statements for a general bivariate normal distribution can be transformed into equivalent statements for a standard bivariate normal distribution with 0 means and variances equal to 1. It is thus sufficient to be able to compute cumulative probabilities for a standard bivariate normal distribution with probability density function

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[x^2 - 2\rho xy + y^2\right]} \quad -\infty < x < \infty \quad -\infty < y < \infty$$

Similar to the case with the univariate normal distribution, the bivariate cumulative distribution function (c.d.f.)

$$F(h,k) = P[X \le h, Y \le k] = \int_{-\infty}^{k} \int_{-\infty}^{h} f(x,y)\ dx\ dy$$

does not have a closed form solution and numerical integration is the usual approach to evaluating such integrals. Because of the importance of the distribution, the National Bureau of Standards

(NBS) has published an extensive set of tables for P[X>h,Y>k] [5]. Another set of tables using a different approximation, has been generated by Owen [6]. Other ways to approximate the bivariate normal integral are discussed in [1], [3] and [7]. All of these approximations involve numerical integration and require extensive computer programming, rendering them to be not readily suitable for obtaining on the spot results. These days, with the easy availability of microcomputers, it would be useful to have a program to compute bivariate normal probabilities interactively and also be able to incorporate such a program within other application programs. This will allow an analyst to perform sensitivity studies by varying the input parameters in an application program and observing the effect on the measures of effectiveness of interest.

Recently, Wang [8] developed a new algorithm to compute bivariate normal probabilities, that does not require numerical integration, relatively easy to program and provides quite accurate results. Investigations by Wang indicate that the computed probabilities compare very well with those in the NBS tables and the computer resources needed are not excessive. The approach used by Wang is to start with an approximate contingency table of cell probabilities for a rectangular grid on the x, y plane and then apply the "iterative proportional fitting algorithm" [3; sec 3.5] to modify the cell probabilities; the iteration is continued until the marginal probabilities in the contingency table coincide with their true values to within a specified degree of accuracy. Since the marginal distributions of the random variables X and Y are univariate normal, the exact

marginal probabilities can be determined using computer programs available in most statistical software packages or by writing a subprogram for the purpose. It should be noted that, although a bivariate normal distribution is defined over the entire plane, for all practical purposes it is sufficient to consider only the square subregion [-4,4] x [-4,4] since the probability content outside of this subregion is negligible.

Wang's algorithm is defined by the following steps:

1. Let $-4 = a_0 \quad a_1 \quad \ldots \quad a_m = 4$ be a partition of the interval [-4,4] along the x - axis and $-4 = b_0 \quad b_1 \quad \ldots \quad b_n = 4$, a partition along the y - axis; identical equal length partitions for both axes is recommended to reduce computational complexity.

2. Let

$$\Delta x = x_i - a_{i-1} \qquad i = 1, 2, \ldots, m$$

$$\Delta y = b_j - b_{j-1} \qquad j = 1, 2, \ldots, n$$

$$m_i = a_{i-1} + a_i \qquad i = 1,2\ldots,m$$

$$n_j = b_{j-1} + b_j \qquad j = 1,2,\ldots,n$$

$$\bar{\rho} = \rho \quad (1 - \frac{\Delta x^2}{12}) \quad (1 - \frac{\Delta y^2}{12})$$

where is the correlation coefficient (specified).

3. Let $\bar{p}_i = p\ [a_{i-1} < x \leq a_i] \quad i = 1,2,\ldots,m$

$\bar{p}_j = P\ [b_{j-1} < Y \leq b_j] \qquad j = 1,2,\ldots,n$

be the marginal probability contents of the i-th and j-th subintervals along the x and y axes respectively.

4. Compute $p_{ij}^{(0)}$, the starting approximate probability of the (i, j) th cell as

$$p_{ij}^{(0)} = e^{\frac{\bar{\rho}\, m_i n_j}{1-\bar{\rho}^2}} \qquad i = 1,2,\ldots m;\ j = 1,2,\ldots,n$$

5. The application of the iterative proportional fitting algorithm results in the following equations for the modified probability of the (i,j) th cell after the k-th iteration:

$$p_{ij}^{(k)} = \begin{cases} \dfrac{p_{ij}^{(k-1)}\, \bar{p}_i}{\sum_{j=1}^{n} p_{ij}^{(k-1)}} & k = 1,3,5,\ldots \\[2ex] \dfrac{p_{ij}^{(k-1)}\, \bar{p}_j}{\sum_{i=1}^{m} p_{ij}^{(k-1)}} & k = 2,4,6,\ldots \end{cases}$$

6. Continue the iteration process until for some even number k

$$\left| \sum_{j=1}^{n} p_{ij}^{(k)} - \bar{p}_i \right| < \varepsilon \quad \text{and} \quad \left| \sum_{i=1}^{m} p_{ij}^{(k)} - \bar{p}_j \right| < \varepsilon$$

where ε is a prespecified degree of accuracy with which the true marginal probabilities agree with the marginals in the contingency table.

7. To compute $P[X \le h, Y \le k]$ (or $P[X > h, Y > k]$) sum the probabilities in the contingency table over those cells for which $a_i \le h$ ($a_i > h$) and $b_j \le k$ ($b_j > k$). In those cases where either $h \ne a_i$ and/or $k \ne b_j$ for any of the partition points a_i and b_j, the

accuracy of the approximation can be improved by including h and/or k as additional partition points.

An APL function, called BVN, to compute bivariate normal probabilities (both $P[X \le h, Y \le k]$ and $P[X > h, Y > k]$) is presented in the appendix. This function invokes another APL function called NCDF to compute the marginal univariate normal probabilities. The BVN function can be run on an IBM-PC / AT compatible microcomputer using an APL language system such as APL*PLUS from the STSC corporation; the function can also be run under VSAPL on the NPS mainframe computer. The function runs interactively and calls for keyboard input of the desired equal length partition size for the x and y axes and the degree of accuracy ϵ in approximating the marginal cell probabilities. With only minor modifications, the function can be imbedded within another APL function as a subprogram. Computations using the BVN function indicate that with partition size $x = y = .2$ for both the x and y axes and $\epsilon = .00005$, a four decimal place accuracy as compared with the NBS tables can be achieved. The computational time, as is to be expected, increases with a decrease in the partition size x or y, an increase in and to a lesser degree a decrease in ϵ. With $x = y = .2$, $\epsilon = .00005$ and $.1 \le \quad \le .8$ the computational time (clock time) was between 90 and 180 seconds on the Zenith Z-248 (an AT type) microcomputer, and on the IBM 3033 mainframe computer these times were between 1 and 20 seconds. The computational time can be reduced considerably (to about 30 seconds on the Z-248) by choosing $x = y = .5$ but then only a three decimal computational

accuracy can be expected. For a fixed value of if the c.d.f. is to be calculated for several choices of (h,k), the BVN function needs to be run only once; with a very minor modification the contingency table of bivariate cell probabilities can be saved in a matrix and all that is left is to sum the probabilities of the appropriate cells. If needed, it would be quite straight forward to generate tables for various choices of and (h,k).

Professor I. O'muircheartaigh of the O.R. department and I are in the process of completing a Fortran program for the problem and expect to submit the code for publication.

REFERENCES

[1] Anscombe, F.J. (1981). Computing in Statistical Science through APL, Springer-Verlag, New York.

[2] Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice, M.I.T. Press, Cambridge, MA.

[3] Gupta, S.S. (1963). "Probability Integrals of Multivariate Normal and Multivariate t", Annals of Mathematical Statistics, 34, 792-828.

[4] National Bureau of Standards (1959). Tables of the Bivariate Normal Distribution Function and Related Functions, Applied Mathematics Series 50, U.S. Government Printing Office, Washington D.C.

[5] National Bureau of Standards (1964). Handbook of Mathematical Functions, Abramowitz, M. & Stegen, I.A. eds., Applied Mathematics Series 55, U.S. Government Printing Office, Washington D. C.

[6] Owen, D.B. (1956). "Tables for Computing Bivariate Normal Probabilities", Annals of Mathematical Statistics, 27, 1075-90.

[7] Sowden, R.R. & Ashford, J.R. (1969). "Computation of Bivariate Normal Integrals", Applied Statistics, 18, 169-80.

[8] Wang, Y.J. (1987). "The Probability Integrals of Bivariate Normal Distributions: A Contingency Table Approach, Biometrika, 74, 1, 185-90.

APPENDIX

THIS APPENDIX CONTAINS THE LISTING OF TWO APL VARIABLES BVNHOW AND NCDFHOW AND TWO APL FUNCTIONS BVN AND NCDF. THE TWO HOW VARIABLES PROVIDE SHORT DESCRIPTIONS OF THE COMPUTATIONAL SCHEMES, THE INPUT PARAMETERS AND THE SYNTAX FOR THE TWO FUNCTIONS.

BVNHOW

THE FUNCTION BVN COMPUTES THE C.D.F. OF A STANDARD BIVARIATE NORMAL DISTRIBUTION WITH CORRELATION COEFFICIENT ρ, USING AN ALGORITHM PROPOSED BY YUCHUNG J. WANG (BIOMETRIKA, 1987, NO.74, 185-90). THE ALGORITHM CONSISTS OF PARTITIONING THE X-Y PLANE INTO RECTANGULAR CELLS, AN INITIAL APPROXIMATION OF THE CELL PROBABILITIES AND AN ITERATIVE SCHEME TO MODIFY THE CELL PROBABILITIES. THE ITERATION PROCESS IS TERMINATED AS SOON AS THE MARGINAL PROBABILITIES COINCIDE WITH THEIR EXACT VALUES (THAT ARE UNIVARIATE STANDARD NORMAL PROBABILITIES,COMPUTABLE USING THE APL FUNCTION NCDF) TO WITHIN A SPECIFIED DEGREE OF ACCURACY ϵ . THE SYNTAX FOR THE FUNCTION IS

RHO BVN W

WHERE RHO IS THE CORRELATION COEFFICIENT AND W = (x,y) IS THE POINT AT WHICH THE C.D.F. IS TO BE COMPUTED. THE FUNCTION WILL CALL FOR THE INPUT OF THE DESIRED LENGTH FOR AN EQUAL PARTITIONING OF THE INTERVAL [¯4,4] ALONG THE X AND Y AXES AND ϵ THE DESIRED DEGREE OF ACCURACY IN THE MARGINAL PROBABILITIES. THE RECOMMENDED CHOICES ARE .2 FOR THE PARTITION LENGTH Δx, AND .00005 FOR ϵ . THE TOTAL COMPUTATION TIME, ON THE ZENITH Z-248 MICROCOMPUTER, SHOULD BE BETWEEN 90 AND 180 SECONDS DEPENDING ON THE SIZE OF RHO.

NCDFHOW

THE FUNCTION NCDF COMPUTES THE C.D.F. OF A STANDARD NORMAL DISTRIBUTION, USING THE APPROXIMATION DEFINED IN EQUATION 26.2.17, PAGE 932 IN THE HANDBOOK OF MATHEMATICAL FUNCTIONS, M.ABRAMOWITZ AND I.A.STEGUN EDITORS, PUBLISHED BY THE NATIONAL BUREAU OF STANDARDS (REF. [5]). THIS APPROXIMATION IS ACCURATE TO ATLEAST TO 7 DECIMAL PLACES. THE SYNTAX FOR THE FUNCTION IS: NCDF Z WHERE Z IS AN INCREASING ARRAY OF NUMBERS FOR WHICH THE C.D.F. IS TO BE COMPUTED. THIS FUNCTION IS INVOKED BY THE BVN FUNCTION TO COMPUTE MARGINAL PROBABILITIES.

```
    ∇ RHO BVN W;X;Y;RBAR;R;A;M;B;E;Q;D;I;J;K;L1;L2;A1;A2;M1;M2;P1;P2
[1]   ' '
[2]   ' '
[3]  ⍝THIS FUNCTION COMPUTES THE CUMULATIVE PROBABILITIES  OF A STANDARD
[4]  ⍝BIVARIATE NORMAL DISTRIBUTION ( means 0 and st.devs 1) WITH
[5]  ⍝CORRELATION COEFFICIENT RHO.  THE SYNTAX FOR THE FUNCTION IS:
[6]  ⍝RHO BVN W  ,  WHERE  W=(x,y).  THE FUNCTION FIRST COMPUTES A
[7]  ⍝CONTINGENCY TABLE OF CELL PROBABILITIES OVER A USER DEFINED PARTITION
[8]  ⍝OF THE X AND Y AXES AND THEN CUMULATES THE PROBAILITIES IN THE
[9]  ⍝APPROPRIATE CELLS.  THE USER IS PROMPTED TO INPUT THE LENGTH OF THE
[10] ⍝SUBINTERVALS IN THE PARTITION OF (¯4,4)  (e.g., .05) AND THE DESIRED
[11] ⍝ ACCURACY (e.g., .0005 FOR A 3-DECIMAL ACCURACY).  THE FUNCTION
[12] ⍝OUTPUTS BOTH Pr [ X < x , Y < y ] AND Pr [ X > x , Y > y ].
[13]  ' INPUT THE DESIRED PARTITION SUBINTERVAL LENGTH FOR X AND Y AXES'
[14]  K←⎕
[15]  ' '
[16]  'INPUT THE DESIRED ACCURACY OF COMPUTATIONS'
[17]  E←⎕
[18]  X←W[1]
[19]  Y←W[2]
[20]  →END1×⍳(X≤¯4)∨Y≤¯4
[21]  →END2×⍳(X≥4)∧Y≥4
[22]  →END3×⍳(X≥4)∨Y≥4
[23]  X←⌊/4,X
[24]  Y←⌊/4,Y
[25]  RBAR←RHO×1-(K*2)÷12
[26]  R←RBAR÷(1-RBAR*2)
[27]  A←¯4+0,K×⍳8÷K
[28]  A1←((X>A)/A),X,(X<A)/A
[29]  A2←((Y>A)/A),Y,(Y<A)/A
[30]  M1←((1↓A1)+¯1↓A1)÷2
[31]  M2←((1↓A2)+¯1↓A2)÷2
[32]  L1←⍴M1
[33]  L2←⍴M2
[34]  B←*R×M1∘.×M2
[35]  P1←(NCDF 1↓A1)-NCDF ¯1↓A1
[36]  P2←(NCDF 1↓A2)-NCDF ¯1↓A2
[37] REPEAT:Q←P1÷+/B
[38]  D←((L1,L1)⍴Q)×((L1,L1)⍴1,L1⍴0)
[39]  B←D+.×B
[40]  Q←P2÷+⌿B
```

```
[41]  D←((L2,L2)⍴Q)×((L2,L2)⍴1,L2⍴0)
[42]  B←B+.×D
[43]  →REPEAT×⍳((+/(|P1−+/B)>E)>0)∨(+/(|P2−+⌿B)>E)>0
[44]  I←+/X>A1
[45]  J←+/Y>A2
[46]  'Pr [ X < ',(⍕X),' ,  Y < ',(⍕Y),' ]  =   ',⍕+/+/(I,J)↑B
[47]  'Pr [ X > ',(⍕X),' ,  Y > ',(⍕Y),' ]  =   ',⍕+/+/(I,J)↓B
[48]  →0
[49] END1:' Pr [ X < ',(⍕X),' , Y < ',(⍕Y),' ]   =  0'
[50]  ' Pr [ X > ',(⍕X),' , Y > ',(⍕Y),' ]   =  1'
[51]  →0
[52] END2:' Pr [ X < ',(⍕X),' , Y < ',(⍕Y),' ]   =  1'
[53]  ' Pr [ X > ',(⍕X),' , Y > ',(⍕Y),' ]   =  0'
[54]  →0
[55] END3:' Pr [ X < ',(⍕X),' , Y < ',(⍕Y),' ]   =   ',⍕(NCDF X)×NCDF Y
[56]  ' Pr [ X > ',(⍕X),' , Y > ',(⍕Y),' ]   =   ',⍕(1−NCDF X)×(1−NCDF Y)
[57]
    ∇
```

```
    ∇ R←NCDF Z;B;p;z;T;N;M;P
[1]   ⍝THIS FUNCTION COMPUTES THE C.D.F. Pr [ Z ≤ z ] OF A STANDARD
[2]   ⍝NORMAL DISTRIBUTION USING THE APPROXIMATION IN EQUATION 26.2.17,
[3]   ⍝PAGE 932 IN THE HANDBOOK OF MATHEMATICAL FUNCTIONS, EDITED BY
[4]   ⍝ABRAMOWITZ AND STATGUN, NATIONAL BUREAU OF STANDARDS.
[5]    Z←,Z
[6]    Z←Z[⍋Z]
[7]    N←⍴(Z<0)/Z
[8]    M←⍴(Z≥0)/Z
[9]    Z←|Z
[10]   p←0.2316419
[11]   T←÷1+p×Z
[12]   z←(*-(Z*2)÷2)÷(○2)*0.5
[13]   B← 0.31938153 ¯0.356563782 1.781477937 ¯1.821255978 1.330274429
[14]   P←1-z×(T∘.*⍳5)+.×B
[15]   R←(1-N↑P),(-M)↑P
    ∇
```

INITIAL DISTRIBUTION LIST

DIRECTOR (2)
DEFENSE TECH. INFORMATION
CENTER, CAMERON STATION
ALEXANDRIA, VA 22314

DIRECTOR OF RESEARCH ADMIN.
CODE 012
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

LIBRARY (2)
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

Center for Naval Analysis
4401 Ford Ave.
Alexandria, VA 22302-0268

DEPARTMENT OF MATHEMATICS
CODE 53
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

PROF.TOKE JAYACHANDRAN (25)
CODE 53Jy
DEPARTMENT OF MATHEMATICS
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

END
DATE
FILMED
JAN
1988